AMENDMENT TO THE CLAIMS

1. Canceled
2. Canceled
3. Canceled
4. Canceled
5. Canceled
6. Canceled
7. Canceled
8. Canceled
9. Canceled

10. (Currently Amended)A method of segmenting a textual input string including characters separated by spaces, comprising:

receiving the textual input string;

proposing a first segmentation of at least a portion of the input string by segmenting the input string at the spaces to obtain a plurality of tokens;

attempting to validate word boundaries in the first segmentation by submitting the first segmentation to a linguistic knowledge component; ~~and~~

if the first segmentation is not validated, proposing a subsequent segmentation by:

determining whether invalid tokens contain any of a predetermined plurality of multi-character punctuation strings or emoticons;

if so, segmenting the tokens into subtokens based on the multi-character punctuation strings or emoticons;

determining whether invalid tokens contain punctuation marks;

if so, segmenting the tokens into subtokens according to a predetermined precedence hierarchy of punctuation;

determining whether invalid tokens contain both alpha and numeric characters;

if so, segmenting the tokens into subtokens at boundaries between the alpha and numeric characters in the tokens;

~~and~~

submitting the subsequent segmentation to the linguistic knowledge component for validation~~.~~; and

repeating the steps of proposing a subsequent segmentation and submitting the subsequent segmentation to the linguistic knowledge component until the portion of the input string is validated or the portion of the input string has been segmented according to a predetermined number of segmentation criteria.

11. Canceled
12. Canceled
13. Canceled
14. Canceled
15. Canceled

16. (Currently Amended) The method of claim ~~15~~ 10 wherein proposing a subsequent segmentation comprises:

reassembling previously segmented subtokens.

17. (Currently Amended) The method of claim ~~11~~ 10 wherein proposing a first segmentation comprises:

identifying a token as a group of characters flanked by spaces or either end of the input string.

18. (Original) The method of claim 17 wherein proposing a

subsequent segmentation comprises:

determining whether the token contains either all alpha characters or all numeric characters; and

if so, indicating that the token cannot be validated.

19. (Original)The method of claim 18 wherein proposing a subsequent segmentation comprises:

determining whether the token includes final punctuation; and

if so, segmenting the token into a subtoken by splitting off the final punctuation.

20. (Original)The method of claim 19 wherein proposing a subsequent segmentation comprises:

determining whether the token includes both alpha and numeric characters; and

if so, segmenting the token into subtokens at a boundary between the alpha and numeric characters.

21. (Original)The method of claim 20 wherein proposing a subsequent segmentation comprises:

determining whether the token includes one or more of a predetermined set of multi-punctuation characters or emoticons; and

if so, segmenting the token into subtokens based on the multi-punctuation characters or emoticons included in the token.

22. (Original)The method of claim 21 wherein proposing a subsequent segmentation comprises:

determining whether the token includes one or more edge punctuation marks; and

if so, segmenting the token into subtokens by splitting off the one or more edge punctuation marks according to a

predetermined edge punctuation precedence hierarchy.

23. (Previously Amended) The method of claim 22 wherein proposing a subsequent segmentation comprises:

determining whether the token includes one or more internal punctuation marks, internal to the tokens; and

if so, segmenting the token into subtokens based on the one or more internal punctuation marks according to a predetermined internal punctuation precedence hierarchy.

24. (New) A method of segmenting a textual input string including characters separated by spaces, comprising:

receiving the textual input string;

proposing a first segmentation of at least a portion of the input string by identifying a token as a group of characters flanked by white spaces or either end of the input string;

attempting to validate word boundaries in the first segmentation by submitting the first segmentation to a linguistic knowledge component;

if the first segmentation is not validated, proposing a subsequent segmentation by:

determining whether invalid tokens contain any of a predetermined plurality of multi-character punctuation strings or emoticons;

if so, segmenting the tokens into subtokens based on the multi-character punctuation strings or emoticons;

determining whether invalid tokens contain punctuation marks;

if so, segmenting the tokens into subtokens according to a predetermined precedence hierarchy of

punctuation;

determining whether invalid tokens contain both alpha and numeric characters;

if so, segmenting the tokens into subtokens at boundaries between the alpha and numeric characters in the tokens;

submitting the subsequent segmentation to the linguistic knowledge component for validation; and

repeating the steps of proposing a subsequent segmentation and submitting the subsequent segmentation to the linguistic knowledge component until the portion of the input string is validated or the portion of the input string has been segmented according to a predetermined number of segmentation criteria.

25. (New) The method of claim 24 wherein proposing a subsequent segmentation comprises:

determining whether the token includes one or more edge punctuation marks; and

if so, segmenting the token into subtokens by splitting off the one or more edge punctuation marks according to a predetermined edge punctuation precedence hierarchy.

26. (New) The method of claim 25 wherein proposing a subsequent segmentation comprises:

determining whether the token includes one or more internal punctuation marks, internal to the tokens; and

if so, segmenting the token into subtokens based on the one or more internal punctuation marks according to a predetermined internal punctuation precedence hierarchy.